

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FACULTAD DE INGENIERÍA MECÁNICA Y ELÉCTRICA

SUBDIRECCIÓN ACADÉMICA



MODELADO Y VISUALIZACIÓN DE RELACIONES
ENTRE CONTAMINANTES DEL AIRE Y SALUD
PÚBLICA

POR

SELENE BERENICE PRADO PRADO

COMO REQUISITO PARCIAL PARA OBTENER EL GRADO DE
INGENIERÍA EN TECNOLOGÍA DE SOFTWARE

JULIO 2022

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FACULTAD DE INGENIERÍA MECÁNICA Y ELÉCTRICA

SUBDIRECCIÓN ACADÉMICA



MODELADO Y VISUALIZACIÓN DE RELACIONES
ENTRE CONTAMINANTES DEL AIRE Y SALUD
PÚBLICA

POR

SELENE BERENICE PRADO PRADO

COMO REQUISITO PARCIAL PARA OBTENER EL GRADO DE
INGENIERÍA EN TECNOLOGÍA DE SOFTWARE

JULIO 2022

Universidad Autónoma de Nuevo León
Facultad de Ingeniería Mecánica y Eléctrica
Subdirección Académica

Los miembros del Comité de Tesis recomendamos que la Tesis «Modelado y visualización de relaciones entre contaminantes del aire y salud pública», realizada por el alumno Selene Berenice Prado Prado, con número de matrícula 1810042, sea aceptada para su defensa como requisito parcial para obtener el grado de Ingeniería en Tecnología de Software.

El Comité de Tesis

Dra. Satu Elisa Schaeffer

Asesora

Dra. Sara Elena Garza Villarreal

Coasesora

Dr. José Arturo Berrones Santos

Revisor

Vo. Bo.

Dr. Fernando Banda Muñoz

Subdirección Académica

San Nicolás de los Garza, Nuevo León, julio 2022

ÍNDICE GENERAL

Agradecimientos	VIII
Resumen	IX
1. Introducción	1
1.1. Motivación	3
1.2. Hipótesis	3
1.3. Objetivos	3
1.3.1. Objetivo general	3
1.3.2. Objetivos específicos	4
2. Antecedentes	5
2.1. Monitoreo de calidad del aire	5
2.1.1. Monitoreo de la calidad del aire en Nuevo León, México	7
2.2. Series de tiempo	7
2.2.1. Clasificación de enfermedades	8
2.3. Modelos de regresión lineal	8

2.3.1. Regresión múltiple	8
3. Estado del arte	9
3.1. Trabajos relacionados	10
3.2. Comparación de trabajos	10
3.2.1. Comparaciones	10
3.2.2. Áreas de oportunidad	10
4. Solución propuesta	11
4.1. Datos recolectados	11
5. Desarrollo de la solución	12
6. Experimentos	13
6.1. Diseño experimental	13
6.2. Resultados	13
6.3. Discusión	13
7. Conclusiones	15
7.1. Contribuciones	16
7.2. Trabajo a futuro	16

ÍNDICE DE FIGURAS

1.1. Localización de las estaciones de monitoreo de la calidad del aire. . .	2
------------------------------------------------------------------------------	---

ÍNDICE DE CUADROS

6.1. Especificaciones técnicas del equipo de cómputo	14
----------------------------------------------------------------	----

AGRADECIMIENTOS

Quiero agradecer a la Dra. Elisa por el apoyo durante el desarrollo de mi tesis y por la motivación y conocimientos brindados para seguir desarrollandome profesionalmente en lo que me gusta. Al programa PAICYT-UANL por su contribución brindada bajo las claves CE1421-20 y CE1842-21.

A mis padres, Lilia Prado López y Adan Alfaro Lerma, por su apoyo y motivación constante desde siempre. A mis hermanos Angel, Estrella, y Adali, a quienes he visto crecer y de quienes he aprendido mucho.

RESUMEN

Selene Berenice Prado Prado.

Candidato para obtener el grado de Ingeniería en Tecnología de Software.

Universidad Autónoma de Nuevo León.

Facultad de Ingeniería Mecánica y Eléctrica.

Título del estudio: MODELADO Y VISUALIZACIÓN DE RELACIONES ENTRE CONTAMINANTES DEL AIRE Y SALUD PÚBLICA.

Número de páginas: 18.

OBJETIVOS Y MÉTODO DE ESTUDIO: El objetivo de la investigación es generar modelos que permitan visualizar relaciones entre contaminantes atmosféricos y salud pública. Los modelos generados se utilizan en conjunto con datos obtenidos de la Secretaría de Salud del Gobierno de México y registros de los niveles de los contaminantes presentes en el área metropolitana de Monterrey.

El tener un modelo que permita visualizar relaciones entre contaminantes atmosféricos y salud pública que sea utilizado con datos confiables y verídicos pueden ayudar a visualizar el impacto que tiene el aumento del nivel de contaminantes atmosféricos.

CONTRIBUCIONES Y CONCLUSIONES: Durante la investigación...

Firma de la asesora: _____

Dra. Satu Elisa Schaeffer

Firma de la coasesora: _____

Dra. Sara Elena Garza Villarreal

CAPÍTULO 1

INTRODUCCIÓN

El crear modelos para la visualización de datos ayuda a observar con mayor claridad los datos para encontrar relaciones entre ellos.

El *aprendizaje máquina*¹ es un área dentro de la *ciencia de datos*² que puede ayudar a crear dichos modelos para tener una más eficiente visualización cuando se trabaja con una gran cantidad de datos, que es lo que se requiere para el presente trabajo. El área de la ciencia de datos es muy útil ya que permite trabajar con grandes cantidades de datos aminorando la cantidad de tiempo empleado en la creación de gráficos que permitan visualizar los datos.

La tarea en el presente proyecto es utilizar modelos para visualizar las relaciones entre los contaminantes del aire y salud pública, para ello se requieren datos sobre salud pública y sobre los niveles de contaminantes del aire.

Para la realización de los experimentos se tienen datos de ingresos hospitalarios provenientes de la base de datos de la Secretaría de Salud del Gobierno de México [7]. También se tienen registros de los niveles de algunos contaminantes del aire presentes en el área metropolitana de Monterrey, dichos registros son hechos por las

¹Traducido como *machine learning* en inglés, tiene como objetivo desarrollar técnicas que les permitan a las computadoras aprender.

²Traducido como *data science* en inglés, involucra métodos para extraer conocimiento de datos, eso con la finalidad de que haya un mejor entendimiento de los datos.

Figura 1.1: Localización de las estaciones de monitoreo de la calidad del aire.

1.1 MOTIVACIÓN

Existen investigaciones que ya han estudiado las relaciones entre contaminantes del aire y salud pública, sin embargo, con el presente trabajo se busca aportar a la creación de nuevas herramientas que permitan observar y estudiar dichas relaciones. El poder visualizar dichas relaciones puede ayudar a tomar medidas adecuadas que permitan aminorar los efectos negativos de los contaminantes del aire en la salud.

1.2 HIPÓTESIS

Se plantea que con modelos de regresión se pueden obtener gráficos donde se pueden observar las relaciones entre el número de ingresos hospitalarios y los niveles de contaminantes del aire.

1.3 OBJETIVOS

En esta sección se establece el objetivo general y los objetivos específicos sobre los que se orienta el presente trabajo.

1.3.1 OBJETIVO GENERAL

El objetivo de generar, implementar y evaluar modelos que muestran las relaciones existentes entre contaminantes del aire y salud pública tiene la finalidad de apoyar a la implementación de estrategias que aminoran los efectos negativos de los contaminantes del aire en la salud de las personas. Con los modelos generados se puede tener una herramienta que permite identificar gráficamente las relaciones con solo proporcionarle el conjunto de datos.

1.3.2 OBJETIVOS ESPECÍFICOS

- Generar, implementar y evaluar modelos de regresión que permite cuantificar las relaciones entre contaminantes del aire y salud pública a partir de un conjunto de datos.
- Diseñar e implementar visualizaciones interactivas que permiten explorar los modelos implementados y su validez estadística.
- Evaluar la eficacia de los modelos generados para que así al utilizar cualquiera de los modelos generados se pueda tener noción de la fiabilidad del análisis realizado a partir de los resultados producidos por los modelos.

CAPÍTULO 2

ANTECEDENTES

Existen factores ambientales que afectan la salud de una comunidad como: el abastecimiento de agua potable y el saneamiento, la vivienda y el hábitat, la alimentación, la contaminación ambiental, el empleo de productos químicos y los riesgos ocupacionales [4].

Contaminación del aire es un término usado para describir la presencia de uno o más contaminantes en la atmósfera, cuyas cantidades y características pueden resultar perjudiciales o interferir con la salud, el bienestar u otros procesos ambientales naturales [5].

En el presente capítulo se presentan los fundamentos y definiciones de los conceptos más relevantes para el tema de estudio abordado.

2.1 MONITOREO DE CALIDAD DEL AIRE

Existen diversos estudios que muestran que existen potenciales efectos a la salud cuando en el aire están presentes contaminantes en forma de partículas, gases o agentes biológicos.

Korc and Sáenz [9] mencionan que desde inicios de 1950 se observa una preocupación por los contaminantes del aire en los países de América Latina y el Caribe. Las universidades y dependencias de los ministerios de salud fueron los organismos que realizaron las primeras mediciones de contaminación en el aire.

En 1965, el Consejo Directivo de la Organización Panamericana de la Salud (OPS) recomendó el establecer programas de investigación de la contaminación del agua y del aire, con el objetivo de colaborar en el desarrollo de políticas adecuadas de control [8].

Mediante el Centro Panamericano de Ingeniería Sanitaria y Ciencias del Ambiente (CEPIS), la OPS acordó establecer una red de estaciones de muestreo de la contaminación del aire. En junio de 1967 La Red Panamericana de Muestreo Normalizado de la Contaminación del Aire (REDPANAIRES) inició sus operaciones recolectando muestras mensuales de polvo sedimentable (PS) y muestras diarias de partículas totales en suspensión (PTS) y de SO₂. La REDPANAIRES comenzó con ocho estaciones y a fines de 1973 tenía un total de 88 estaciones distribuidas en 26 ciudades de 14 países [8].

Para diciembre de 1973 se habían recolectado más de 350,000 datos sobre la calidad del aire, en los que se observa que algunas ciudades mostraban una tendencia al incremento de los niveles de contaminación [8].

En 1980 la REDPANAIRES desapareció y pasó a formar parte del Programa Global de Monitoreo de la Calidad del Aire, iniciado en 1976 por la OMS y el Programa de las Naciones Unidas para el Medio Ambiente (PNUMA), como parte de un sistema global de monitoreo ambiental llamado GEMS por sus siglas en inglés *Global Environmental Monitoring System*.

En la década de 1990, la OMS organizó, con carácter global, el Sistema de Información para el Control de la Calidad del Aire llamado AMIS por sus siglas en inglés *Air Management Information System*. Entre las actividades más destacadas de AMIS se incluye el coordinar las bases de datos sobre temas relacionados con la

calidad del aire.

2.1.1 MONITOREO DE LA CALIDAD DEL AIRE EN NUEVO LEÓN, MÉXICO

En Nuevo León, México, las operaciones de la Red Automática de Monitoreo Atmosférico iniciaron en 1993. Dicha red en sus inicios contaba con cinco estaciones fijas de monitoreo continuo de monóxido de carbono (CO), dióxido de azufre (SO₂), óxidos de nitrógeno (NO_x), ozono y PM₁₀ [9]. Como se muestra en la figura 1.1, actualmente se cuenta con nueve estaciones fijas.

2.2 SERIES DE TIEMPO

Korc and Sáenz [9] mencionan que las relaciones entre niveles de concentraciones de contaminantes del aire y los efectos sobre la salud generalmente son obtenidas de estudios epidemiológicos de series de tiempo. Uno de los diseños epidemiológicos más utilizados son los estudios de series temporales. Con esos diseños se analizan las variaciones en el tiempo de la exposición al contaminante y el indicador de salud estudiado en una población [2].

Las series de tiempo se pueden definir como un conjunto de observaciones ot tomadas en un tiempo t determinado. Los estudios de series de tiempo relacionan estadísticamente los cambios temporales en la repercusión de cambios en la concentración de un contaminante en la población [3].

Para mostrar datos en una serie de tiempo, especialmente en el área médica, estos suelen agruparse en *semanas epidemiológicas*¹.

¹Una semana epidemiológica es un estándar de medición temporal que se utiliza para comparar datos en ventanas de tiempo definidas. La primera semana epidemiológica del año termina el primer sábado de enero de cada año [1]

2.2.1 CLASIFICACIÓN DE ENFERMEDADES

Existe un instrumento estadístico y sanitario para identificar enfermedades llamado Clasificación Internacional de Enfermedades (CIE), cuya finalidad es entender las causas de morbilidad y mortalidad de la población y así mejorar la calidad de vida de la misma. Es en base a un criterio epidemiológico y sanitario establecido por Farr a finales del siglo XIX que esta clasificación agrupa enfermedades en epidémicas, generales, locales ordenadas por origen geográfico, trastornos del desarrollo y lesiones [10]. Para lograr distinguirlas se emplea un código alfanumérico que consiste de una letra en la primera posición, seguida de dos dígitos, un punto decimal y un último dígito. El rango de valores va de A00.0 a Z99.9.

2.3 MODELOS DE REGRESIÓN LINEAL

La tendencia w_0 de una serie de tiempo puede ser obtenida a partir de una regresión lineal de la misma [6]. Una regresión lineal es una metodología inferencial supervisada que busca predecir valores y dado un vector de variables de entrada t por medio del ajuste de coeficientes w de la función lineal

$$\hat{y}(t, w) = w_0 + w_1x_1 + \dots + w_tx_t. \quad (2.1)$$

2.3.1 REGRESIÓN MÚLTIPLE

Un modelo de regresión múltiple es un modelo complemento de la regresión lineal simple, el cual tiene dos o más variables independientes k que pueden influir en una variable dependiente y . Peniche-Camps and Cortez-Huerta [11] expresan la regresión múltiple mediante la siguiente ecuación:

$$y = \beta_0 + \beta_1x_1 + \dots + \beta_kx_k + \varepsilon. \quad (2.2)$$

CAPÍTULO 3

ESTADO DEL ARTE

En el presente capítulo se estudia literatura reciente relacionada con el presente trabajo, esto con el objetivo de revisar distintos métodos para resolver el problema planteado en el presente trabajo y, además, también revisar implementaciones similares para resolver problemas distintos. Lo anterior tiene la finalidad de comparar los trabajos revisados e identificar áreas de oportunidad en ellos.

En la primera sección, trabajos relacionados, se recopilan obras con características relacionadas al presente trabajo, ya sean relacionados con el problema que se pretende resolver o con los métodos empleados para buscar su resolución.

En la segunda sección, análisis comparativo, se comparan las distintas características de los trabajos revisados, de esa forma se pueden determinar las principales ventajas y desventajas de cada trabajo.

Finalmente, en la tercera sección, áreas de oportunidad, se realiza una conclusión acerca de los resultados obtenidos del análisis comparativo.

3.1 TRABAJOS RELACIONADOS

3.2 COMPARACIÓN DE TRABAJOS

La mayoría de los trabajos citados...

3.2.1 COMPARACIONES

En el cuadro...

3.2.2 ÁREAS DE OPORTUNIDAD

En el cuadro...

CAPÍTULO 4

SOLUCIÓN PROPUESTA

Habiendo conocido las características que mejor describen a los atributos del presente trabajo, se puede decir que la base del método propuesto se puede desarrollar...

4.1 DATOS RECOLECTADOS

Inicialmente...

CAPÍTULO 5

DESARROLLO DE LA SOLUCIÓN

Recapitulando las fases anteriores, se conoce que...

CAPÍTULO 6

EXPERIMENTOS

Después de...

6.1 DISEÑO EXPERIMENTAL

Hola...

6.2 RESULTADOS

Establecidos los experimentos que se van a realizar, se reporta los resultados obtenidos...

6.3 DISCUSIÓN

Todos los experimentos son ejecutados en una laptop con las especificaciones del cuadro 6.1.

Cuadro 6.1: Especificaciones técnicas del equipo de cómputo

Sistema Operativo	Windows 10 64 bits
Procesador	Intel Core i5-7300HQ
RAM	8 GB RAM DDR4 2133 MHz

CAPÍTULO 7

CONCLUSIONES

Este capítulo describe la tesis a partir de la manera que cumple los objetivos generales y específicos para determinar si la hipótesis se comprueba, trata también del porque se realizó la tesis...

7.1 CONTRIBUCIONES

La solución propuesta surgió a partir de...

7.2 TRABAJO A FUTURO

La solución propuesta en la tesis...

BIBLIOGRAFÍA

- [1] ARIAS, J. R. (2006), «What is an epidemiological week and why do we use them», *Skeeter*, **66**(1), page 7.
- [2] BALLESTER DÍEZ, F., J. M. TENÍAS and S. PÉREZ-HOYOS (1999), «Efectos de la contaminación atmosférica sobre la salud: una introducción», *Revista Española de Salud Pública*, **73**(2), pages 109–121.
- [3] BROCKWELL, P. J., P. J. BROCKWELL, R. A. DAVIS and R. A. DAVIS (2016), *Introduction to time series and forecasting*, Springer.
- [4] CATALÁ, F. and E. DE MANUEL (1998), «Informe SESPAS 1998: La salud pública y el futuro del estado del bienestar», *Granada: EASP*.
- [5] CORBITT, R. and R. CORBITT (1990), *Standard Handbook of Environmental Engineering*, McGraw-Hill.
- [6] DARLINGTON, R. B. and A. F. HAYES (2016), *Regression analysis and linear models: Concepts, applications, and implementation*, Guilford Publications.
- [7] DIRECCIÓN GENERAL DE INFORMACIÓN EN SALUD (2021), «Egresos Hospitalarios», URL http://www.dgis.salud.gob.mx/contenidos/basesdedatos/da_egresoshosp_gobmx.html.
- [8] HADDAD, R. (1974), «Contaminación del aire. Situación actual en la América Latina y el Caribe», *Informe técnico*. <http://www.bvsde.ops-oms.org/bvsacd/scan/004499/004499-02B.pdf>.

-
- [9] KORC, M. and R. SÁENZ (1999), «Monitoreo de la calidad del aire en América Latina», *Korc Marcelo E*, pages 1–22.
- [10] ORGANIZATION, W. H. *et al.* (2016), «International Classification of Diseases. 2016», *URL: <http://www.who.int/classifications/icd/en/>*[accessed 2018-09-27].
- [11] PENICHE-CAMPS, S. and M. CORTEZ-HUERTA (2020), «La costumbre al envenenamiento: El caso de los contaminantes atmosféricos de la ciudad de Guadalajara, México», *Revista de Ciencias Ambientales*, **54**(2), pages 1–19.
- [12] SIMA (2015), «Sistema Integral de Monitoreo Ambiental», *URL <http://aire.n1.gob.mx>*.

RESUMEN AUTOBIOGRÁFICO

Selene Berenice Prado Prado

Candidato para obtener el grado de
Ingeniería en Tecnología de Software

Universidad Autónoma de Nuevo León
Facultad de Ingeniería Mecánica y Eléctrica

Tesis:

MODELADO Y VISUALIZACIÓN DE RELACIONES ENTRE
CONTAMINANTES DEL AIRE Y SALUD PÚBLICA

Nací el 30 de Junio de 2000 en Monterrey, Nuevo León, soy la mayor de cuatro hijos. Mi familia está conformada por mi madre Lilia Prado López, mi padre Adan Alfaro Lerma, y mis hermanos: Angel Alejandro Prado Prado, Estrella Belen Prado Prado, y Genesis Adali Alfaro Prado.

Desde pequeña me han gustado las matemáticas, aprender como funcionan los sistemas computacionales, y leer.

Durante los primeros semestres de mi carrera descubrí la inteligencia computacional, un área que me encantó desde que la descubrí, en especial su rama de ciencia de datos, rama en la que espero seguir desarrollándome.

Otra cosa que me apasiona es dibujar y pintar, actividades que estaban dentro de mi pero que se avivaron cuando inició la pandemia en el año 2020.